International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 5, Issue 2, Apr 2015, 89-96

AN ADVANCED RESEARCH FRAMEWORK ON DENTAL HEALTH CARE INFORMATION ECO-SYSTEM FOR ENHANCED SERVICES BY LEVERAGING BIG DATA ANALYTICS

V. V. R. MAHESWARA RAO[1] & V. VALLI KUMARI[2]

[1]Professor, Department of Computer Science & Engineering, Shri Vishnu Engineering College for Women, Andhra Pradesh, India

[2]Professor, Department of Computer Science & Systems Engineering, Andhra University, Andhra Pradesh, India

ABSTRACT

The data in dental health domain is rich as it is collected from different sources like public survey records, electronic health records, genomic data and behavioral data and so on. In addition, the insight of this data is complex, noise, heterogeneous, longitudinal, high volume and incremental in nature. As a result, effective handle and make inferences from such dental health care data by conventional techniques were proved to be inefficient and thus, necessitate the newest research towards the Big Data Analytics. The big data analytics is in its infancy for the dental health care domain and is evolving into a promising right research direction. It also refers to the innovative process of collecting, organizing and analyzing high volume, high velocity, high variety and high veracity information assets that discover knowledge for enhanced insight and decision making in evidence based dental health care.

Handling this big data in dental health domain faces a series of technical challenges, where scalable, adoptive and robust approaches are needed. To address such challenges, the authors in the present paper present "Dental Health Care Information Eco System (DHCIES)" that comprehensively and equally concentrates on all stages of big data analytics. Initially in the infrastructure stage, the system focuses on required physical infrastructure facilities. Subsequently in next stage, it emphasizes on the techniques of data acquisition, integration, and computation of data. Finally in the application stage, it explores big data analytic functions including statistical analysis, clustering, and classification and so on. The eco system provides a solid foundation for diagnosis of dental health care and creates increased awareness of the importance on the dental health care to overall health. It also reduces disparities in accessing to dental health care. This promotes improved treatment plan and advancing health policy reform.

KEYWORDS: Big Data, Dental Health Care, Big Data Analytics, Data Acquisition, Data Management, Statistical Analysis, Clustering, Classification

INTRODUCTION

The rapid advancements in digitization, the amount of health care data being generated from various resources have reached astronomical proportions. This data is structured, semi structured, unstructured, complex, heterogeneous, high dimensional and incremental in nature. At the same time, patients are increasingly demanding information about their dental health care options, so that, they understand their choices and can participate in decisions about their health care. Extracting useful real time information on such enormous health care data has evolved as solid base for the present researchers to enter into the era of big data analytics.

Big data not only means a fundamental shift in the way data is stored and managed, it also entitles deploying powerful real time analytics and visualization tools, collaboration platforms and the ability to automatically create links with the existing applications such as business support systems and customer relationship management.

In technology perspective, big data is the possibility of better storage -Volume, the ability to process the information and make it available in real time - Velocity and the ability to deal with various kinds of data sources, including structured, semi-structured and unstructured ones - Variety. Inclusion of Veracity as the fourth big data attribute emphasizes the importance of addressing and managing for the uncertainty inherent with in some times of data. The technology exists, so the essential issue is how carriers can make sense of the massive volumes of data and deliver value to business.

Fundamentally, big data means not only 4V's of data but also describes a new generations of technologies and architectures, design to economically extract value from very large volumes of a wide variety of data, by enabling velocity capture, discovery and analysis. This definition converge the four dimensions: Volume, Variety, Velocity and Veracity as shown in figure 1 help both to define and discribe big data.

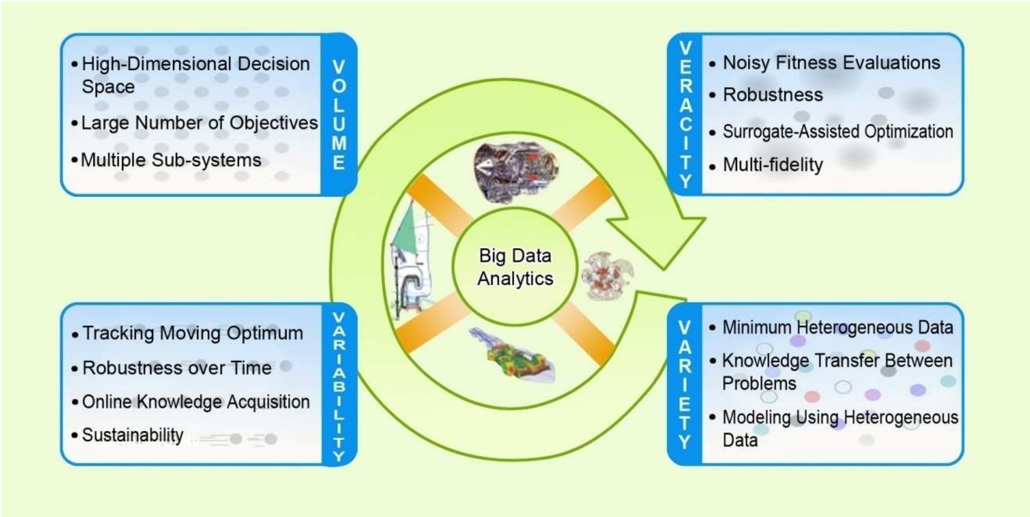

Figure 1: Characteristics of Big Data

Volume: Volume indicates the size of the data, the volume of big data evolved into its present stage as mega bytes to giga bytes, giga bytes to terra bytes, terra bytes to peta bytes, peta bytes to exa bytes. The big data volumes are relative and vary by factors, such as time and the type of data. This data is too big to be handled by the current state of techniques and systems. In future, this will continue to expand exponentially at an unprecedented rate, is a prime motivation to create revolutionary data management mechanisms.

Velocity: Velocity refers to the rate at which data are generated and the speed at which it should be analyzed and acted upon. Velocity in big data is a concept which processed and analyzed with the speed of the data coming from various sources. The proliferation of digital devices such as smart phones and sensors has led to an unprecedented rate of data which is continually being generated at a pace that is impossible for traditional systems to capture, store and analyze. This, coupled with the need and drive to be more agile and deliver insight quicker, is a basic motivation for the present data engineering researchers to build the necessary infrastructure and skill base to react quickly.

Variety: Variety shows different types of data and data sources. Variety in big data is a measure of heterogeneity of data representation such as structured, semi structured and unstructured. With the explosion of sensors, smart devices and social collaboration technologies, data is being generated in countless forms, including: text, web data, tweets, sensor data, audio, video, click streams, log files and more. However, the emergence of new data resources enables the researchers towards new management technologies and analytics, which enable to leverage data in an innovative aspect.

Veracity: Veracity denotes data uncertainty. Veracity in big data is the level of reliability associated with certain types of data. Some data is inherently uncertain, for example: sentiment, truthfulness, weather conditions, economic factors. The need to acknowledge and plan for this dimension of uncertainty of big data is still a major quality concern in processing of data. This triggered the present researchers towards the robust optimization techniques to manage uncertainty.

Thus, the authors in the present paper propose "Dental Health Care Information Eco System (DHCIES)" that builds on the amalgamation all characteristics of big data analytics. At first, the system pays an attention on infrastructure stage, to recognize possible scalable network, storage and security infrastructure. Further, it processes on data acquisition, integration and computation in computational stage. Lastly, the system performs big data analytic functions including statistical analysis, clustering, and classification and so on.

This paper is organized as follows. Initially, the authors describe related work. The next section, explores the proposed work. Finally, conclusions and future work are made.

LITURATURE SURVEY

The research work in this paper contains the literature survey from 2010 to the current year for each stage of health care big data analytics.

In the year 2010, [18] wrote a report on the promise and peril of big data. They provided the basic concepts of big data and explained business and social implications of big data. They have expressed that medical researchers focus on to identify useful correlations between medical treatments and health outcomes, that helps to improve health and medical care can be made more efficient and effective.

In the subsequent year 2011, [15] addressed three key technologies for extracting real-time business value from the big data. They summarized that only one-third of organizations do big data analytics, the present research given the newness of the combination of advanced analytics and big data.

All the range in 2012, [14] the IBM institute developed a report based on analysis of survey data and discussions with academics, subject experts, business executives. They mainly concentrated on the concept, characteristics, sources, analytics capabilities, adoption stages and primary obstacles of big data. In their conclusion, they clearly suggested that future research demands the effective use of information and analytics to understand comprehensive needs of the organizations.

In the same year, [13] explained the evaluation changes of digital world and their causes in 21^{st} century. Later they presented the big data pipeline and list out the challenges in big data. They finally described that big data analytics as an emerging type of knowledge work, with plenty of opportunities in different domains.

In the year 2013, they [12] made an attempt on use of big data for Communication Service Providers (CSP) to take important decisions and activities such as designing more competitive offers, prices and packages. They presented big data analytics stack and flexible layered analytics platform for CSR. In addition, they revised the opinion of big data and presented the new view of big data. They concluded that the researchers have to pay attention not only on the characteristics of big data should also work on analytics of big data.

In 2014, Xindong Wu, Xingquan Zhu et.al., [3] have given the evidence based literature on how big data applications have grown tremendously. They presented a HACE theorem to explain features of the big data and also its framework. In addition, they proposed a big data processing model in data mining perspective. Finally, they list-out the challenges of big data analytics and need of big data mining in all science & engineering domains.

Recently in the year 2015, Amir Gandomi, Murtaza Haider [1] have attempted to offer a broader definition of big data and its characteristics. They primarily focused on the analytic methods to leverage massive volumes of heterogeneous data in unstructured text, audio and video formats. However, they worthily emphasized that real world adoption of big data analytics were not economically feasible for large scale applications. Also, in their conclusion, they felt that novel big data analytics not yet taken place and it becomes a prolific field of research.

PROPOSED WORK

To understand the complete significance of the knowledge embedded in dental health data, one must recognize the innumerable forms of multifaceted associations that emerge when data is positioned in a much broader context of overall dental health care. The main challenges in handling dental health data lie not only in huge Volume in amount, high Variety in type, Velocity in terms of real-time requirements, and Variability, but also in the approach to understanding data.

To create effective analytics that produce actionable intelligence for dental health data, the authors in the present paper proposes a three stage "Dental Health Care Information Eco System (DHCIES)" is shown in figure 2 that comprehensively and equally concentrates on all stages of big data analytics in bottom up approach. Initially, in the big data infrastructure stage, the system focuses on network infrastructure, storage infrastructure and security infrastructure facilities. Subsequently in next stage, it emphasizes on the techniques of data acquisition, integration, management and computation of data. Finally in the application stage, it explores big data analytic functions including statistical analysis, clustering, and classification and so on.

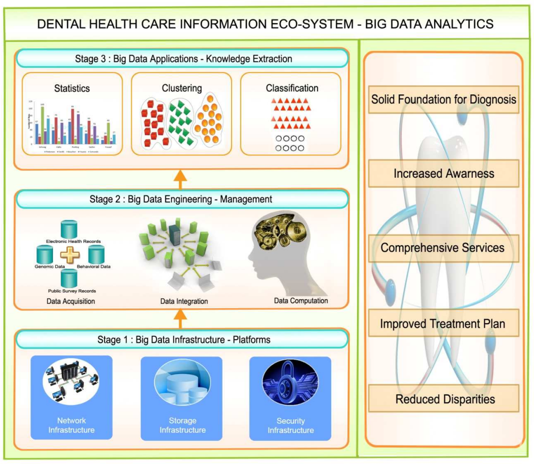

Figure 2: Architecture of Dental Health Care Information Eco System (DHCIES)

Big Data Infrastructure

To achieve significant, measurable knowledge from big data can only be realized if scalable infrastructure put into place as a solid foundation that support rapidly growing volume, variety, velocity and veracity of data. With this aim, the initial stage of proposed DHCIES recognizes possible scalable hardware infrastructure with high capacity storage warehouse is the most prevalent foundation component in the environment of big data analytics, each sustains the rapid growth of current and future data. The big data infrastructure mainly consists of a collection of scalable network and high capacity storage resources along with security infrastructure.

Today, dental health data exist in more forms than ever before, creating an extraordinary challenge. To better serve patients demands for information everywhere, dental health centers must develop new strategies for optimizing multiple kinds of networks that have flexibility and agility. In fact, network infrastructure cited more often than the establishment of the analytics platforms themselves. It is able to tune instantaneously to address various real world demands and different types of application environments dynamically.

The nature of big data, the certainty of its size and the analysis and workflows it must sustain, puts a lot of burden on the storage infrastructure as well. Capacity and performance efficiency must be maintained in order to keep the costs of storing and handling such large amounts of data under control. One of the challenges that big data brings is the requirement to support many different data types, or at least have that ability. Big Data’s volume may rapidly outgrow existing storage, causing purchasing organizations to look for affordable capacity wherever they can. This can lead to the acquisition of storage systems from different manufacturers and a need to combine the capacity on diverse platforms.

A big data infrastructure should provide data security assurance so that this investment is appropriately cared for. Obviously, a traditional backup process can be impractical, since making weekly, or even daily backup copies of large numbers of very large files could take too long. Techniques such as attribute based encryption may be necessary to protect sensitive data and apply access controls. In addition, the security requirements have closely aligned to specific needs.

This stage is exposed to upper stages in a superior grained way to execute a specific service. In this stage infrastructural resources are allocated to produce the actionable intelligence on the health care data to meet the targeted users.

Big Data Engineering & Management

Big data engineering is a core component of any analytics effort and it is even more important, yet much more complex, with big data. In order to address rightly, the second stage of proposed system encapsulates various data tools that runs over raw data recourses. In addition, data management refers to mechanisms and tools that provide persistent data storage and highly efficient management, such as distributed file systems and SQL or NoSQL data stores. In this context the proposed system uses data sources & acquisition, data integration and data computation/querying methods.

Data Sources & Acquisition: This stage of proposed DHCIES collects the data in large, diverse and complex datasets from various longitudinal and distributed data sources, including electronic health records, public survey records, behavioral data, genomic data and other available digital sources. In general, these datasets are associated with diverse levels of domain specific values. In addition, there are also technical challenges in collecting and processing, these datasets. Moreover, the collected datasets contain many worthless data, which unreasonably increases the amount of storage space and influences further process.

Data Integration: Data integration means acquiring data from disparate sources and integrating the dataset into a unified form with the necessary data pre-processing operations. Generally, one of the fundamental characteristics of the dental health care data is the huge volume of data represented by varied and different dimensionalities. This is because the patient information such as health records, X-ray images, family health history, genomic data and so on are recorded in their own schemas. Later, the integrated data need a high-speed transmission method to transmit the data into the proper storage supporting system for a variety of queries in analytical applications.

Data Computation: To apply any big data technology on the high volume data it is necessary to implement any agile data driven and data exploration techniques that works with application logic and facilitates the big data analytics in any application smoothly. In general, any big data technique works completely in parallel manner. This technique reduces the data from large data set based on a function. In addition, to analyze or interact with the stored data, storage systems must provide several interface functions, fast querying and other computing models.

The query driven data is fed to application stage as an input. The big data application stage extracts the knowledge by applying various big data analytic that produces the actionable intelligence on the health care data to meet the target users.

Big Data Applications – Knowledge Extraction

The big data applications layer extracts the interface provided by the data computing models to implement various data analysis functions, including, statistical analyses, clustering, and classification.

Statistical Analysis: The science of the collection, organization, and interpretation of data, including the design of surveys and experiments. Statistical techniques are often used to make judgments about what relationships between variables could have occurred by chance and what relationships between variables likely result from some kind of underlying causal relationship. Towards this, the proposed DHCIES, initially creates desired survey forms by implementing the theories and techniques of statistical approach to collect the dental health data in a suitable form. Later, the system derives statistical values based on disease types, population wise, poverty lines, gender wise, region wise and so on.

Classification Analysis: A set of techniques to identify the categories in which new data points belong, based on a training set containing data points that have already been categorized. One application is the prediction of segment specific customer behavior where there is a clear hypothesis or objective outcome. These techniques are often described as supervised learning because of the existence of a training set. With this aim, the DHCIES applies advanced classification analysis on patient dental health data to identify patient similarity based on training dataset.

Cluster Analysis: A statistical method for classifying objects that splits a diverse group into smaller groups of similar objects, whose characteristics of similarity are not known in advance. An example of cluster analysis is segmenting consumers into self-similar groups for targeted marketing. This is a type of unsupervised learning because training data are not used. In order to segment the dental health data the proposed DHCIES uses cluster methods based on characteristics of patient, diseases, behavior and genomic data.

OUTCOMES OF PROPOSED WORK

The proposed DHCIES is worked over a period of patient's dental health care data under standard execution

environment. The system comprehensively and equally concentrated on all stages of big data analytics on the given dental health care data. The proposed system ensures the prediction in better and more scientific patient classification and clustering within reasonable amount of time.

The outcomes of DHCIES show oral health and dental visiting patterns of children, adolescents, adults and seniors and trends over a period. In addition, the reports clearly indicates the financial barriers, services received, water fluoridation, treatment needs of various patient groups. Moreover, the proposed DHCIES is guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies, thus garner the big data researcher's attention.

The DHCIES is compared with the traditional techniques in terms of performance. The experimental results indicate a noticeable improvement of DHCIES performance over the traditional techniques.

CONCLUSIONS

The proposed Dental Health Care Information Eco System yields immediate returns in terms of patient outcomes and lowering care costs by efficiently utilizing the colossal dental health care data repositories. In addition to that, the model provides right intervention to the right patient at the right time in a sensible manner. This system not only helps the dentists to understand the information contained within the data, but it also ensures to identify the data that is most important to the future real time predictions in dental domain. This system is evidence that every dental health care organizations need to devote time and resources to understanding this phenomenon and realizing the envisioned benefits. Finally, big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions.

REFERENCES

1. Amir Gandomi, Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management, Vol. 35, pp: 137–144, 2015.
2. Han Hu, Yonggang Wen,Tat-Seng Chua, Xuelong Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", IEEE Transactions and Content Mining, Vol. 2, pp: 652-687, 2014.
3. Xindong Wu, Xingquan Zhu,Gong-Qing Wu, Wei Ding,"Data Mining with Big Data", ieee transactions on knowledge and data engineering, vol. 26, no. 1, pp: 97-107, 2014.
4. Bellagio Big Data Workshop Participants, "Big data and positive social change in the developing world", Rockefeller Foundation Bellagio Centre conference, pp: 1-35, 2014.
5. Craig C. Douglas, "An Open Framework for Dynamic Big-Data-Driven Application Systems (DBDDAS) Development", 14th International Conference on Computational Science, Vol. 29, pp:1246–1255, 2014.
6. Sergey V. Kovalchuk, Artem V. Zakharchuk, Jiaqi Liao, Sergey V. Ivanov, Alexander V. Boukhanovsky, "A Technology for BigData Analysis Task Description using Domain-Specific Languages", 14th International Conference on Computational Science, Vol. 29, pp: 488–498, 2014.
7. Yang Xu, Mingming Zeng, Quanhui Liu, and Xiaofeng Wang, "A Genetic Algorithm Based Multilevel Association Rules Mining for Big Datasets", Hindawi Publishing Corporation Mathematical Problems in

Engineering, pp: 1-9, 2014.

8. Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, Graham J. Williams, “Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives”, IEEE Computational intelligence magazine, pp: 62-74, 2014.
9. Omar El-Gayar, Prem Timsina, “Opportunities for Business Intelligence and Big Data Analytics In Evidence Based Medicine”, 47th Hawaii International Conference on System Science, pp: 749-757, 2014.
10. Phillip A. Laplante, “Who’s Afraid of Big Data?”, IEEE Computer Society, pp:6-7, 2013.
11. A TCS Global Trend Study, “The Emerging Big Returns on Big Data”, Tata Consultancy Services, 2013.
12. Ericsson White paper, “Big Data Analytics”, 284 23-3211 Uen | August 2013.
13. Danyel Fisher, Rob DeLine, Mary Czerwinski, Steven Drucker, “Interactions with Big Data Analytics”, Coverstory, interactions, 2012.
14. IBM Institute, “Analytics: The real-world use of big data How innovative enterprises extract value from uncertain data”, IBM Global Business Services Business Analytics and Optimization, Executive Report, 2012.
15. SaS, “Big Data Meets Big Data Analytics”, White Paper, 2011.
16. Steve LaValle, Eric Lesser, Rebecca Shockley,et.al., “Big Data, Analytics and Path from Insights to Value”, MIT Sloan Management Review, 2011.
17. Brad Brown, Michael Chui, James Manyika, “Are you ready for the era of ‘big data’?”, mckinsey global institute, 2011.
18. David Bollier Rapporteur, “The Promise and Peril of Big Data” The Aspen Institute, ISBN: 0-89843-516-1, 2010